MEDICAL LIBRARY

HUMAN BIOLOGY

a record of research

SEPTEMBER, 1946

Vol. 18 No 3

PUBLISHED QUARTERLY BY
THE JOHNS HOPKINS PRESS, BALTIMORE, MARYLAND, U.S.A.

HUMAN BIOLOGY

a record of research

FOUNDED BY RAYMOND PEARL

Edited by

CHARLES P. WINSOR

in association with

A. W. FREEMAN H. S. JENNINGS ADOLF MEYER
CURT P. RICHTER ADOLPH H. SCHULTZ

Managing Editor

FRANCES D. NICOLLS

ADVISORY BOARD

ED. CLAPARÈDE
University of Geneva

S. R. DETWILER
Columbia University

EUGEN FISCHER
Kaiser Wilhelm Institut für Anthropologie, Berlin-Dahlem

CORRADO GINI
R. Università di Roma

MAJOR GREENWOOD
School of Hygiene and Tropical Medicine, London

SHINKISHI HATAI
Tôhoku Imperial University, Sendai

ERNST KRETSCHMER
Universitäts-Nervenklinik, Marburg

A. J. LOTKA
Metropolitan Life Insurance Company

D. PEYRONY
Musée Préhistorique, Les Eyzies (Dordogne)

P. RIVET
Universidad de Bogotá, Colombia

JOHN B. WATSON
New York

CLARK WISSLER
American Museum of Natural History

HUMAN BIOLOGY is published quarterly by The Johns Hopkins Press, Baltimore, Maryland, U. S. A. The advance subscription price is $5.00 a year (foreign postage 35 cents; Canadian 13 cents). Single current issues are $2.00 each; back numbers, $3.00 each; back volumes, $6.00; plus postage. Subscriptions for less than a year are charged at the single issue price.

Subscribers should notify the publishers of change in address at least two weeks before publication of the issue with which change is to take effect. No claim for non-receipt of an issue will be entertained unless made immediately upon receipt of the next succeeding issue.

Manuscripts and communications regarding editorial matters should be addressed to Editor, Human Biology, 615 N. Wolfe St., Baltimore—5, Maryland, U. S. A.

THE JOHNS HOPKINS PRESS, *Publishers*

BALTIMORE—18, MARYLAND, U. S. A.

Entered as Second Class matter, February 18, 1929, at the Post Office at Baltimore, Md., under the Act of March 3, 1879.

Copyright 1947 by The Johns Hopkins Press

HUMAN BIOLOGY
a record of research

SEPTEMBER, 1946

VOL. 18 No. 3

FACTORIAL ANALYSIS OF ANTHROPOLOGICAL MEASUREMENTS IN PSYCHOTIC PATIENTS

BY T. V. MOORE, O.S.B., M.D., PH.D. AND E. H. HSÜ, PH.D.
The Catholic University of America

IN AN attempt to determine empirically the associations of psychotic symptoms into syndromes, the senior author of this study associated with him in the work a distinguished physical anthropologist, Dr. C. J. Connolly, and asked him to undertake certain anthropological measurements of the patients studied. The object of this was to test Kretschmer's theory of the association of body build and psychotic condition in a material in which the syndromes were empirically determined and anthropological measurements were made by one who had a high degree of professional skill. A very remarkable association was found and the results published by Dr. Connolly (11).

The problem then arose of considering the anthropological measurements as a body of empirical data interesting in themselves and without necessary relations to the psychotic conditions. We have, however, taken up the relation of anthropological measurements to psychotic type as revealed by the method of this study. The data were intercorrelated by Dr. M. Gertrude Reiman, at that time statistical assistant in our Department of Psychology.

The next stage was a factorial analysis of the table of intercorrelations. Dr. Reiman's table is given below (See Table 1).

Since many readers may not be familiar with factorial analysis it may be helpful to prelude this study with an historical note, the purpose of which is to give an insight into the meaning and importance of the analysis of measurements into groups of one or more factors.

Factorial analysis had its origin in psychology. In the early part of this century psychologists commenced to apply the now well-known Bravais-Pearson correlation formula[1] to psychological measurements. Thorndike of Columbia became very much interested in the study of mental life by the intercorrelations of tests. But he had no technique for factorial analysis and came to what Spearman later called the anarchical concept of the mental life as group performances underlying which there were no general abilities of any kind.

The first glimmering of factor analysis appeared in an article by Spearman entitled "General Intelligence Objectively Determined and Measured" and published in the *American Journal of Psychology* in 1904 (Vol. 15, pp. 201-293). Spearman had long looked upon general intelligence as a special mental function underlying all performances. The basis for this concept was that he developed various techniques for demonstrating that for the most varied groups of tests it was possible to find a general underlying factor that could account mathematically for all the intercorrelations of the tests. The last and best known of his techniques was what is known as the method of tetrad differences.

If we intercorrelate n tests and have also the correlation of each of the n tests with a criterion or one underlying general factor, we can write n equations such as the following:

$$a_1 + a_2 r_{12} + \ldots\ldots a_n r_{1n} = r_{10}$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$a_n r_{1n} + a_2 r_{2n} + \ldots\ldots a_n = r_{n0}$$

The tetrad difference criterion is satisfied when all the first order determinants of these equations (containing four correlation coefficients involving any four variables) reduce to zero within the limits of probable error. In this case this set of equations derive from a group of equations in which there is only one linearly independent variable.

It was not long before the universality of Spearman's G was questioned and one commenced to find criteria for the presence of two

[1] $r = \frac{1}{N} \frac{\Sigma xy}{\sigma_x \sigma_y}$

or more *g*-factors, and attempts to analyze correlation equations for *n*-factors, whether group factors or general factors (3, 17, 31, 35).

What is probably the most complete technique for such an analysis was given by Louis Leon Thurstone in his classic book *The Vectors of Mind* (28). Thurstone's technique has been used in the present study.

Since this technique and its implications may not be widely known among biologists, it may be well to say something concerning the nature of the factors that are isolated by this and similar methods.

There are two extreme views of the meaning of the factors isolated. One is that of Thurstone:

> Factor analysis is reminiscent of faculty psychology. It is true that the object of factor analysis is to discover the mental faculties. (30)

The other extreme is that of Cyril Burt:

> We use factors in psychology as we use rectangular coordinates in other sciences, not because we believe that the phenomena investigated are necessarily dependent upon a few, isolated, independent causal agencies, which operate in the mind or brain, and which a sound method of factor analysis should successfully isolate, but merely because such simplified descriptions enable us to organize our facts into a more logical system and help us to state our inductive arguments more cogently, and to endow our predictions with higher probabilities. (3)

However, it has been repeatedly found that the correlations of two or more tests designed to measure identical mental functions are higher than those between tests that measure disparate functions. This is evident in the early data of Clark Wissler (34).[2] What clouded this result was the tendency of many investigators to correlate all sorts of performances without any effort to design specific tests for specific functions.

Then some forms of factorial analysis leave the investigator in a stage of his work in which sound interpretation and reference to the realities originally measured are impossible. Such, for instance, would be the technique of Thurstone if he stopped with the "centroid method" and did not go on to a rotation of the axes. Such also is the difficulty with Hotelling's method, as Thurstone has pointed out.[3]

Anthropological measurements have an advantage over most psycho-

[2] Findings by Anastasi (1) and Monaghan (21) seem to clarify this point.

[3] (28) pp. 131-2.

logical tests in that the number and size of the errors of measurement are reduced to a minimum.

It seems reasonable too to look upon the size of the measurements as a function of growth. For all organs develop from the realm of the infinitesimal to a value plus or minus the average of the species.

Can we say that growth is growth and that there can be only one growth factor in an organism? All we know about the thyroid and the plurality of substances in pituitary secretion indicates that this question must be answered in the negative. There are a number of growth factors. Furthermore, they are *real,* not arbitrary, ways that we adopt to describe the organism.

How many growth factors are there, and by what anthropological measurements may they be detected? The present study is a contribution towards the solution of this problem. It is presented as a preliminary investigation, not a final and complete study. Had we introduced more measurements we would have had further ways of determining the factors isolated and might well have detected factors for which we have no evidence in the present study.

LITERATURE ON THE PROBLEM

At an early date Spearman attempted to intercorrelate anthropological measurements. Unfortunately, he found no "*g*" in the 7 measures of bodily dimensions from 3,000 male adults. Similar results were obtained by factorial analyses of a battery of 8 developmental and physical traits such as area of ossification, height, weight, chest girth, nutrition, etc., obtained from 115 subjects, and of a battery of 6 anthropological tests obtained from 477 boys and girls. Spearman therefore concluded that divisibility into two factors does not occur in anthropometric data. But we must remember the fact that what Spearman was searching for was a general factor underlying all the measures without any overlap on group factors. If we are interested in the discovery of common factors the data cited by Spearman can be easily factored into several dimensions. For example, there is apparently a group factor comprising the following variables: left midfinger, left cubit, left foot and height. Another group factor can be easily discovered by mere inspection, such as area of ossification, ratio of ossification, height, and nutrition. This seems to be a developmental factor. A third group factor may be discovered from these variables: standing height, sitting height, weight and vital capacity (27).

Cohen (7, 9) applied Thurstone's centroid and Burt's method of least squares to 14 linear measurements made on a group of 64 psychotic males. Two factors through the analysis were identified as isometric and allometric, implying respectively "a common rate of somatic development in all bodily regions" and "growth divergence" which is responsible for the linear or circumferential development. Cohen (8) also measured 50 male students in a factorial study which was in agreement with his earlier result. Burt (2) measured 131 selected and 197 unselected children of school age, 73 selected and 100 unselected adolescents, as well as 186 adults, and correlated the measures of height, weight, circumferences of chest and stomach, length of arm and leg, and length and breadth of skull. The figures were indicative of the presence of three "separate and independent" factors: (1) a well-marked general factor with saturation coefficients all positive, seemingly a factor of growth of all directions; (2) a fairly well-marked bi-polar factor, making, when positive, for disproportionate length of the long bones; (3) a second bi-polar factor, most clearly marked among adults, making, when positive, for a disproportionate increase in breadth, thickness, girth, and weight. Burt suggested three modes of classification accordingly. Regression weights were found for a number of individuals in the selected temperamental groups to measure each of the three tendencies. But the correlations between the physical measures and the temperament factors found previously, were not significant. According to Burt, relative shortness seemed to be correlated to a very small extent with cheerful emotions and relative thinness a little more strongly with inhibitive or regressive tendencies. The correlations seldom rose above .15. After the scheme was reduced to only two antithetical types, i.e., a "plump" type and a lean type, regression equations were worked out for the best weighted expression for "plumpness" and for "general emotionality," "extroversion" and "cheerfulness." The highest correlation, which amounted to —.33, was found between "general emotionality" and "plumpness" in the selected adolescent group. A correlation of .32 on the other hand was found in adult group between the same physical factor and extroversion. Burt concluded that such correlations could not be trusted for the needs of diagnosis. Burt in 1943 (4) applied his new technique of "multiple factor analysis" to the correlation matrix of more than 500 sixteen-year-old American youths obtained from Harvard University and found a general factor accounting for more than 60 per cent of the total variance and a bi-polar factor distinguishing persons showing

relatively excessive growth in "vertical" directions and those showing relatively excessive "transverse" or "circumferential" growth. Burt regarded this confirmation of his earlier findings as an illustration that types are not sharply discriminable classes and that "the more powerful general factor" exerts an obscuring influence on the "types."

Hofstätter (15) made three analyses of 15 subjects which gave rise to three typologies of 5 types each. McCloy (19) used 19 anthropometric scales to measure boys and girls from three months to 16 years. After analyzing the data by Thurstone's methods, he found the following factors: (1) growth in fat, (2) general growth, (3) growth in cross-section, and (4) an unidentified factor. Using Thurstone's methods, Waldrop (33) examined Sheldon's hypothesis which describes the human body build in terms of three major components formulated on the basis of twenty measures selected from a larger number by multiple correlation technique. Mullen analyzed the data obtained from 1320 cases on 17 physical measurements with Holzinger's bi-factor method, and found three factors: a general factor of size and two group factors. "One of these is found in the residual correlations of height, span of arms, length of forearms, and length of lower leg . . . The other group factor is found in the residual correlations for weight, bi-iliac and bi-trochanteric diameter, and chest girth, width and depth." (22) Holzinger and Harman (17) analyzed 8 physical variables of Mullen's anthropometric data for 305 fifteen-year-old girls with principal-factor solution and found three factors, namely, (1) general physical growth, (2) lankiness (height, arm span, length of forearm and length of lower leg), and (3) stockiness (weight, bi-trochanteric diameter, chest girth and chest width).

Sterrett (26) presented an intercorrelation matrix of nine anthropometric variables. A factor having vanishing tetrads was singled out therefrom by us, a factor which is composed of knee breadth, hip breadth, chest breadth and face length. The last one seems to be out of place in this group. Judged from the asymmetry[4] of the figures in the given intercorrelation matrix some of the coefficients are evidently erroneous. Therefore the presence of face length in this factor is very doubtful. Coleman (10) isolated a height factor and a "dead weight" factor from a matrix of intercorrelations of 2 physical measures and 9 tests for physical fitness. Carpenter (5) using Thurstone's technique analyzed the intercorrelations of two physical measurements

[4] In several instances $r_{ij} \neq r_{ji}$

and 6 track and field tests on 100 college girls. He likewise singled out from the matrix a factor identified as dead weight. Carpenter (6) in another study gave an intercorrelation table of ten anthropometric variables. A factor which is composed of neck, forearm, calf, and elbow is apparently present in the male population but not in the female population. Dearborn and Rothney (13) gave an intercorrelation table of 5 physical measures and one mental measure. We were able to discover one physical factor therefrom which can be identified with weight, iliac, chest depth and chest width. This physical factor undoubtedly underlies the data cited. Sanford and his co-workers (25) offered an intercorrelation table of 18 surface and Roentgenographic measurements. They "distinguished" four syndromes of body build, namely, (1) wide, heavy build, (2) tall, narrow build, (3) large build (tall, heavy, and wide) and (4) short, wide build. By inspecting the given matrix, however, one may easily discover several factors, such as a factor including trunk height, suprasternal height, iliac height and head length; a factor of weight, chest breadth, chest depth and chest circumference; a factor of standing height, sitting height, biacromial width and bicristal width. Hammond (16) applied Burt's method of least squares to a 12-variable problem and an 8-variable problem. Hammond made use of the anthropometric data obtained from 100 North Ireland adult males and from 40 individuals taken at random from Wales. He singled out three factors from the first-mentioned sample, namely, (1) the general factor for physical measurement ("It will determine the size which these various organs will attain."); (2) a "body-type factor,"—growth in length and depth and breadth; and (3) a factor consisting of both head and length groups. The Welsh group had 8 variables out of which he found two factors, i.e. the "first general factor" as found in the other sample and a factor composed of the length measure to form one group and head and face the other. Rees (23) in a factorial analysis of 21 physical traits on 200 soldiers found a general factor for the growth considered as "general body size" and a bi-polar factor, associated with length and breadth, respectively, in the growth of human body. Rees and Eysenck (24) measured physically and mentally 389 successive admissions to an emergency hospital together with a control sample of 100 non-neurotic soldiers of similar age distribution and military status, there being 17 measures in all. The intercorrelations of the body measurements were calculated for 200 neurotic soldiers of the 389. This matrix was analyzed by Burt's summation method, which gave rise to two "significant" factors, namely, a body-

size factor and a body-type factor which measures "specific growth in length" as opposed to growth in breadth. Based on the factorial pattern, an index supposedly measuring the body-type factor was derived which is the quotient of 100 times stature divided by transverse chest diameter times six. The distribution of indices of a sample of 400 neurotic soldiers differ from normal by a positive skewness. Those with an index of body-type more than one standard deviation above the mean were called by the authors "leptomorphs," those with an index within plus or minus one standard deviation about the mean, "mesomorphs," and those with an index more than one standard deviation below the mean, "eurymorphs." The "eurymorphs" tend towards "hysteria" and "contain fewer schizoid personalities" than do the "leptomorphs" who tend towards the affective disorders and contain a high percentage of schizoid personalities. Hammond's data were reanalyzed by Thurstone (30) whose result gave a more sensible and clearer-cut structure. Four factors out of the battery of 12 were found and identified as (1) factors that are concerned with the head size, (2) bone length, (3) transverse measures or girth, and (4) size of extremities. Thurstone drew the following conclusion: "The factors that determine body size and proportions seem to constitute a system that is simpler than the factors which determine the cognitive or intellective functions." Thus there seems to be a real difference between the mental measures and the physical data.

McDonough (20) measured 28 boys and 22 girls from grades 7 and 8, with an average age of 170.4 months, in 22 anthropometric ways. She got a pyknotic g-factor which is composed of chest breadth, hip girth, foot breadth, and head breadth, and an asthenic g-factor which has such variables as head length, foot length, face length and crinion and nose length.

Finally one may ask what is the nature of these physical factors. It will be interesting to note that Huxley (18), using methods entirely different from factorial analysis, was able to divide growth into multiplicative, intussusceptive type and additive, accretionary type. He claimed that the normal growth-ratio of a heterogonic organ is to be determined by a result of balance between the size of the organ and its body, as described by the equation $y = bx^k$ (18). Whenever $y < bx^k$ the growth ratio of y is accelerated ($k_1 > k$), and becomes normal again ($k_1 = k$) when the organ reaches the size demanded by the original equation. In a study of the Molluscs, Huxley asserted that body forms depend on constant differential growth-ratios which are of

four "factors": (1) the ratio of growth in length to that in width, (2) the median growth-ratio, (3) the lateral growth-ratio, and (4) the ratio of excessive growth at specific points to that manifested in the major growth-gradients. Thus there seems to be some underlying principle which determines the differential growth-gradients which in turn determine the differential growth-rate of an organ or an organism. This underlying principle might be interpreted as a physical factor.

THE DATA AND THE ANALYSIS

The data concerning the 100 patients have been given by Connolly as already referred to, so in this paper his figures will not be reprinted. Thirty-one variables from the original measures were selected by Dr. M. G. Reiman who was responsible for the preparation of the intercorrelations. No index value was included for index values involve a ratio between two variables. Each variable is a single physical measurement. These thirty-one variables were intercorrelated by the product-moment method. The correlation matrix is presented in Table 1. It was subjected to an analysis (centroid) as developed by Thurstone (28). After the seventh factor was extracted, the standard error for the average seventh residual was found to be .067, while the standard error for a zero correlation should be .100. The standard error for the average first residual was .196. According to Coombs' criterion (12) which makes use of the number of negative signs to detect the presence of common factors, the total number of negative signs appearing in the seventh residual matrix should be 384 after the change of signs while in the present case they are 382. The more negative signs present, the less the vectors come together to form a degenerate cone after the reflections, hence the less the common factor dimensions remain. Tucker's criterion (32) requires the value of .937 to demonstrate that a satisfactory number of common factor dimensions have been reached. In our case the value was found to be .925. This is an empirical criterion which is the ratio of the sum of the absolute values of the correlations, including the diagonals used, before any factor is extracted, to the sum of the absolute values of the residuals after the factor is extracted. This ratio should be less than $(n - 1)/(n + 1)$, where n is the total number of variables.

In spite of these slight discrepancies there are good reasons to believe that a residual factor has been obtained and this is verified indeed by the result of the final rotation. After the seventh dimension

TABLE 1*

Correlation matrix

(Each figure is to be multiplied by 0.01)

VARIABLE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1. Stature																														
2. Sitting height	69																													
3. Span	70	48																												
4. Length of head	13	52	02																											
5. Breadth of head	11	12	17	34																										
6. Height of head	29	40	17	29	40																									
7. Length of face	39	31	34	23	17	24																								
8. Breadth of face	09	23	35	04	45	26	40																							
9. Diameter frontal minimum	16	09	13	07	15	14	08	14																						
10. Diameter bigonial	21	12	27	-01	20	19	26	50	16																					
11. Palpebral length	25	-08	20	00	11	19	25	24	06	25																				
12. Palpebral breadth	12	17	02	-07	00	11	02	-12	44	-14	18																			
13. Interpupillary distance	49	21	27	22	36	40	27	35	23	20	34	02																		
14. Nasion to prosthion	57	37	35	29	21	29	69	33	07	34	30	06	35																	
15. Length of nose	40	20	29	07	08	18	42	27	04	19	06	06	-04	51																
16. Breadth of nose	12	23	18	15	12	23	02	07	06	-01	13	04	22	-02	-06															
17. Breadth of mouth	31	13	18	08	19	16	07	13	10	05	10	02	15	19	36	41														
18. Length of ear	25	25	31	17	16	19	23	15	07	20	11	08	11	19	60	27	14													
19. Breadth of ear	11	08	28	21	23	08	32	53	04	04	-05	04	08	27	35	05	11	44												
20. Ant.-post. diameter of chest	49	19	26	18	24	25	16	27	13	36	00	-02	14	12	09	13	03	28	03											
21. Lateral diameter of chest	25	33	37	27	25	38	31	41	17	34	28	-04	28	22	16	23	12	40	10	54										
22. Circumference of chest	03	38	37	-07	28	42	21	42	30	34	23	-10	25	16	05	20	12	43	02	62	85									
23. Waist girth	19	31	21	22	28	36	13	30	14	26	21	-10	20	07	04	17	00	49	-03	57	76	84								
24. Hip girth	37	36	34	26	25	38	20	31	14	29	20	-09	26	19	11	16	-02	46	03	58	76	82	85							
25. Calf girth	32	45	38	31	16	30	30	10	14	28	22	02	35	31	12	23	08	33	12	42	63	68	62	71						
26. Fore-arm girth	29	50	42	23	09	34	34	37	17	27	30	02	33	34	12	24	16	25	00	43	66	71	58	65	82					
27. Length of hand	71	50	69	11	11	33	41	23	11	23	09	11	16	43	36	11	19	23	26	20	30	30	13	26	27	38				
28. Breadth of hand	48	50	56	24	11	31	34	39	22	32	22	08	32	35	31	29	18	28	12	37	57	58	46	56	63	72	58			
29. Length of arm	76	47	77	08	22	30	37	30	11	29	14	12	32	39	18	06	11	23	10	21	26	30	16	30	25	33	73	47		
30. Circumference of neck	24	37	26	30	25	38	22	32	15	-34	55	02	26	16	06	30	18	33	01	46	54	72	64	67	57	64	23	49	25	
31. Length of leg	83	49	71	01	14	30	38	31	06	11	13	11	16	28	20	11	06	19	-13	02	19	23	09	24	22	26	48	39	70	12

* Figures given in this table have been rounded off from original three digits.

was reached, two of the communalities turned out to be larger than unity. It is believed that the reliability of the physical measurements is very high and hence there is some error involved in taking the highest correlation in the column as the estimated communality. The ideal method is of course to reanalyze the correlation matrix with the communalities reestimated, and this procedure is repeated until a satisfactory stage is reached. But to analyze a 31-variable problem calls for a considerable computational labor. One really wonders what will be gained after the repeated estimation of the communalities and factoring of the matrix, when one considers the tremendous computational work involved. Another cause to account for the larger than unity common factor variance is probably the small sample we employed which was only 100 cases. Such variance was one of the reasons to lead us to stop seeking any higher dimensions for our factorial problem. The 7-dimensional centroid matrix is given in Table 2.

Thurstone's 1938 method (29) of rotating the centroid matrix into a simple structure was employed in this study. After 12 cycles of rotations the transformation matrix was obtained. It is presented in Table 3. When this matrix is pre-multiplied by its transpose it gives the cosines of the angular separation between any two pairs of the reference vectors. This is shown in Table 4.

The rotated matrix, V ($V = F\,\Lambda$)[6] is given in Table 5. The first factor designated by the letter A consists of the following variables (all with a factor loading above 0.4): stature, span, nasion to prosthion (length of face), length of nose, length of ear, length of hand, length of arm, and length of leg. This factor is clearly of a linear character. The grouping seems to be in fair agreement with some of Kretschmer's length measures. It would be interesting to see whether this factor would differentiate the schizophrenics from the cycloid psychotics or not.

The next factor designated by the letter B consists of such variables as sitting height, span, length of face, breadth of face, bigonal diameter, nasion to prosthion, forearm girth, length of hand, breadth of hand, length of arm and length of leg. Both linear and lateral measures came into one group and the picture is marked by the appearance of the "area" measures of both face and hand. This seems to be a general magnitude factor, i.e., bigness in all dimensions. It is thus

[6] Here V is the rotated matrix; F is the centroid matrix; and Λ is the matrix of direction cosines.

tentatively labelled as a factor having something to do with both length and breadth. However, the interpretation of this factor should wait until further analysis is made on a much larger battery of physical measures.

TABLE 2

*Centroid matrix**

VARIABLE	I	II	III	IV	V	VI	VII	h^2
1.	.70	−.54	.23	−.27	−.15	−.07	.35	1.05*
2.	.56	−.33	.43	−.09	−.14	.30	−.34	.84
3.	.68	−.38	.26	.03	−.12	−.21	.11	.74
4.	.27	.24	−.32	−.15	.32	−.27	.34	.55
5.	.40	.14	−.33	.04	−.27	.10	.23	.43
6.	.54	.10	−.14	−.17	−.16	.18	.06	.41
7.	.55	−.27	−.21	.05	.13	−.25	−.17	.53
8.	.54	.07	−.29	.31	−.24	−.20	−.09	.58
9.	.28	.09	−.16	−.18	−.30	.22	−.26	.35
10.	.39	.10	−.19	.32	−.37	−.34	−.18	.59
11.	.36	.13	−.17	−.20	.06	−.12	.06	.24
12.	.10	−.17	−.13	−.33	−.10	.31	−.23	.32
13.	.48	.12	−.22	−.30	−.21	−.16	.13	.48
14.	.57	−.33	−.29	−.06	.11	−.26	−.16	.62
15.	.39	−.47	−.26	.30	.34	.12	−.07	.68
16.	.30	.10	−.07	−.24	.13	.23	.17	.26
17.	.28	−.12	−.24	−.17	.10	.23	.05	.25
18.	.52	−.10	−.10	.37	.32	.31	.14	.65
19.	.27	−.23	−.40	.34	.11	−.05	.08	.42
20.	.52	.29	.19	.25	−.14	.12	.09	.50
21.	.73	.40	.12	.20	.07	.03	−.06	.75
22.	.72	.51	.30	.28	−.08	.22	−.18	1.02
23.	.62	.53	.23	.27	.09	.19	.11	.85
24.	.71	.43	.30	.20	.09	.04	.09	.84
25.	.70	.29	.23	−.08	.27	−.10	−.17	.74
26.	.73	.30	.24	−.16	.19	−.15	−.29	.85
27.	.62	−.44	.19	−.06	−.05	−.12	−.06	.64
28.	.76	.05	.18	−.12	.08	−.12	−.22	.69
29.	.62	−.40	.24	−.08	−.31	−.20	.17	.78
30.	.63	.39	.21	−.26	.36	.24	.19	.89
31.	.51	−.43	.29	−.12	−.26	−.12	.20	.67

* This table has been rounded off from a three-place table from which the calculations were made to a two-place table.

The third factor (C) is composed of the following variables: stature (loading .416), calf girth (loading .524), forearm girth (loading .603), breadth of hand (loading .494) and circumference of neck (loading .603). Considering the three variables highly loaded with this factor, one may easily see that it is a lateral or circumferential factor chiefly of the forearm, calf, and neck. It will be interesting to compare this lateral measure with the linear measure on the different psychotic groups.

Factor D has such variables as breadth of head, (loading .543), height of head (loading .422), breadth of face (loading .472), bigonial diameter (loading .437), anterior-posterior diameter of chest (loading .549), lateral diameter of chest (loading .523), circumference of chest (loading .658), waist girth (loading .604), hip girth (loading .539).

TABLE 3

Transformation matrix, A

	A	B	C	D	E	F	G
I	.33	.50	.36	.46	−.17	.24	.07
II	−.78	−.34	.09	.52	.00	.12	.28
III	−.19	.20	.34	−.13	−.30	−.57	−.66
IV	−.02	.24	−.80	.23	−.37	−.27	.06
V	.26	−.37	.28	−.59	−.40	−.15	.35
VI	.18	−.23	−.08	.19	.61	−.55	−.35
VII	.38	−.59	−.13	.26	−.46	.46	−.48

TABLE 4

Correlations between reference vectors

	A	B	C	D	E	F	G
A	1.00	.02	.00	−.25	−.15	.13	−.23
B	.02	1.00	.02	.10	.05	−.19	.05
C	.00	.02	1.00	−.23	.03	.07	−.03
D	−.25	.10	−.23	1.00	.10	.28	−.13
E	−.15	.05	.03	.10	1.00	−.26	.04
F	.13	−.19	.07	.28	−.26	1.00	.32
G	−.23	.05	−.03	−.13	.04	.32	1.00

This factor seemingly also pertains to breadth or circumference especially from the head to the hip. However, if one takes only those two variables having the highest loadings into consideration, it seems to point towards the circumference of the trunk. This factor is tentatively identified as a factor relative to the thickness of the trunk.

Factor E consists of minimal frontal diameter and palpebral breadth.

TABLE 5

Simple structure

VARIABLE	A	B	C	D	E	F	G
1.	.69	.38	.42	.11	−.23	.27	−.51
2.	.25	.65	.36	.07	.20	−.42	−.34
3.	.44	.55	.24	.14	−.34	.15	−.25
4.	.14	−.31	.20	.11	−.35	.58	.33
5.	.12	.03	−.10	.54	.08	.38	.04
6.	.14	.15	.23	.42	.16	.22	−.00
7.	.36	.44	.13	−.02	−.18	.25	.32
8.	.04	.48	−.19	.47	−.11	.32	.29
9.	−.08	.25	.13	.30	.44	.02	.08
10.	−.15	.52	−.22	.44	−.10	.29	.27
11.	.06	.02	.27	.16	−.06	.33	.20
12.	.14	.11	.25	−.04	.48	−.09	−.01
13.	.08	.12	.29	.37	.02	.52	.12
14.	.42	.43	.20	−.05	−.11	.33	.34
15.	.63	.26	−.14	−.16	−.13	−.04	.19
16.	.18	−.16	.28	.16	.07	.12	−.04
17.	.32	−.03	.15	.06	.16	.12	.07
18.	.46	.09	−.11	.19	−.19	−.08	.03
19.	.39	.14	.37	.08	−.16	.22	.25
20.	−.08	.25	.02	.55	−.16	.00	−.13
21.	−.09	.30	.20	.52	−.22	.06	.14
22.	−.27	.40	.17	.66	−.07	−.20	−.00
23.	−.16	.10	.13	.60	−.25	−.06	−.03
24.	−.10	.22	.25	.54	−.33	.01	−.04
25.	−.06	.29	.52	.18	−.25	.03	.19
26.	−.13	.41	.60	.20	−.18	.04	.21
27.	.46	.56	.29	.01	−.16	.05	−.16
28.	.09	.50	.49	.19	−.14	.06	.12
29.	.42	.54	.24	.24	−.23	.22	−.36
30.	.07	−.14	.60	.29	−.16	.05	−.05
31.	.44	.44	.25	.13	−.20	.14	−.43

The identification of this factor is again to wait for further research.

Factor F reveals three variables with high positive loadings, namely, the length of the head, the breadth of the head, and interpupillary distance. This is clearly indicative of a factor underlying the cephalic measurements and size of the head. It would be of considerable interest to see if the cephalic factor has anything to do with the types of psychoses or psychological make-up.

Factor G reveals no high positive loadings in any variable and only a high negative one which is stature. If the signs are reversed, there is still but one high positive loading. To such a factor it is impossible to assign any meaningful interpretation. This is probably a residual factor, and with the appearance of such a factor, the conclusion that a satisfactory number of common factors has been extracted seems justified.

In summary, by Thurstone's factorial techniques, we have arrived at a simple structure which reveals at least five meaningful factors, i.e., a linear factor (A), a magnitude factor (B), a lateral factor (C), a circumferential factor (D), and a factor that has to do with the size of the head (F).

The next question we want to ask is whether or not these factors are independent or correlated. With this in mind, we have found the intercorrelations between the primary vectors as presented in Table 6. Following Thurstone's formula we got this correlation matrix by premultiplying and post-multiplying the inverse of the matrix of the cosines of angular separation between the reference vectors with a normalizing diagonal matrix, D.[6] From this correlation matrix, it is easily seen that factor A is positively correlated with factor D and negatively correlated with factor F. Factor C is positively correlated with factor D, and factor D is negatively correlated with factor F. The factors that might be considered as independent are A and C; C and F; B, whatever it may be, is relatively unrelated with all the rest except for a slightly high correlation with factor F. Factor E, whatever it may be, is only slightly correlated with factor F. As G is considered a residual factor its correlational significance with the rest is not considered. The moderately high correlation between the length factor and the breadth factor calls for interpretation. At first sight they should be negatively correlated if Kretschmer's theory about the pyknic and leptosome physique is correct. But if one considers the nature of our

[6] See below, Table 6.

TABLE 6

Correlation between primary vectors

	A	B	C	D	E	F	G
Linear factor A	1.00						
Magnitude factor B	−.19	1.00					
Lateral factor C	.17	−.14	1.00				
Circumferential factor D	.45	−.29	.35	1.00			
Palpebral breadth E	−.04	.07	−.14	−.24	1.00		
Size of head F	−.36	.33	−.27	−.57	.34	1.00	
G	.43	−.23	.20	.47	−.19	−.55	1.00

$D\ (A'A)^{-1}D$

sampling which consists of a good number of athletic types in which both the linear and the lateral factors are well developed, one may be easily lead to the conclusion that in our present sampling the athletic type probably has had a considerable share in the total variance. Should the components of our sampling be changed in such a way that it contains only a homogeneous group of leptosome physique, the present correlation between these two factors probably would change into a negative value. The fact that C and D are slightly correlated may be easily explained by the combination of both breadth and athletic factors, if D is identified with athletic physique. Both D and C are correlated or close to each other, as already pointed out by Connolly without a factorial analysis. The fact that factors A, C and D are all negatively correlated with factor F, the cephalic factor, is interesting. It must be borne in mind that our cephalic factor is not a cephalic index. But it has long been pointed out by a number of investigators that the head shape, if expressed in cephalic index or any other index, bears some significance in differentiating one race from another as the somatic type does. Thus the Alpine type tends to have brachycephalic index and also a pyknotic physique. Although the cephalic index may not be related to our cephalic factor, it is nevertheless evident that in some way this cephalic factor is essentially different from the other factors, yet has some antagonistic relationship with them. The highest negative correlation is that between the circumference of the trunk and the cephalic factor. It is, however, necessary to mention that none of the

correlations between the primary factors is really very high. They range from .45 to —.57.

The next question we should ask is: How would these factors found thus far differentiate one type of psychosis from another? Such a step would test the validity of any theory of physical types. Bearing this objective in mind, we have to convert all the raw anthropometric values into standard scores, $(x - m)/\sigma$, multiplied by their respective regression weights. We thus selected 37 non-paranoid schizophrenics, 32 paranoid schizophrenics, and 27 manic-depressives. One case diagnosed as "paranoid condition" was included in the paranoid group. Depressions of various types were predominant in the cyclothymic group. This predominance is perhaps owing to the fact that patients with a manic excitement may be difficult to subject to anthropometric study. It should be pointed out that in the present study more cases in each group have been included than the original series of 100 patients. There were available 10 schizophrenics and 14 manic-depressives that had not been utilized to form the group of 100 to obtain the original intercorrelations. These extra cases were embodied in compiling Table 7.

The next step is to select the best variables to represent each factor. Only the highest factor loadings are considered. Since factor B and factor E are less accessible to meaningful interpretation, they are represented by only one variable each, that which has the highest loading on that factor. Thus for A we selected:

1. Stature
3. Span
14. Nasion to prosthion
15. Length of nose
18. Length of ear
27. Length of hand
29. Length of arm
31. Length of leg;

For factor B we selected:

2. Sitting height;

For factor C:

25. Calf girth
26. Forearm girth
30. Circumference of neck;

For factor D:

22. Circumference of chest[7]
23. Waist girth;

For factor E:

12. Palpebral breadth;

For factor F:

4. Head length
5. Head breadth
13. Interpupillary distance;

For factor G:

7. Length of face.

Then by linear combination the centroid loadings of these variables in each group under the heading of each factor were added up, and the resulting combination together with that of the specifics was normalized. By so doing, we have reduced the rank of an original 31×31 order matrix to that of a 7×7 one. This would save no little computational work in finding the inverse. Then the transformation was performed by post-multiplying this 7×7 matrix with the normalized inverse of the matrix of direction cosines. This transformation gives the correlation between the composite variables and the primary factors. The resulting matrix was transposed and it was pre-multiplied by its own transpose. This resulting multiplication gives the intercorrelations between the composite variables, and is given in Table 8. Then the inverse of this matrix was found. The transpose of the matrix of correlations between the primaries and the composite variables was finally post-multiplied by this inverse matrix and this operation gives the desired regression weights. The formula as derived by Thurstone is

$$w_{pd} = R_{pc} R_{dc}^{-1},$$

where p is primary and c and d stand for composite vectors ($c = d$). The regression weights are given in Table 9. By inspecting this table one will see that except for a high negative weight of factor F on the

[7] It would have been possible to determine this factor by still other measures and to have taken fewer measures for A, but the measures taken are the highest and determine the presence of the factor in a symmetrical matrix.

TABLE 7

Raw scores (anthropometric values) of each psychotic group according to four factors

FACTOR	A								C			D		F			N	AGE	
VARIABLE	1	3	14	15	18	27	29	31	25	26	30	22	23	4	5	13		M	SD
M: TOTAL*	169.8	171.9	7.00	5.3	6.5	18.2	72.8	95.1	34.1	25.5	37.1	89.0	80.6	19.5	15.3	6.2	100		
Non-paranoid schizophrenic	170.6	171.7	7.11	5.4	6.4	18.2	73.1	96.3	32.7	24.8	35.9	85.3	75.5	19.3	15.2	6.2	37	32	8.6
Paranoid schizophrenic	172.4	175.5	7.10	5.4	6.7	18.5	74.2	96.4	34.6	26.8	35.7	93.8	85.2	19.5	15.2	6.3	32	40	11.2
Manic-depressive	168.7	170.7	6.95	5.3	6.7	18.2	72.3	95.3	33.4	25.1	36.6	89.8	84.1	19.4	15.3	6.2	27	50	13.6
σ: ($N = 100$)	6.7	9.0	0.56	0.38	0.49	1.03	3.72	4.84	3.12	2.16	2.73	7.58	9.13	0.84	0.76	0.31			

* Connolly's data

TABLE 8

Intercorrelations between the composite variables (R_{ca})

	C_A	C_B	C_C	C_D	C_E	C_F	C_G
C_A	1.00						
C_B	.47	1.00					
C_C	.26	.26	1.00				
C_D	.11	.23	.33	1.00			
C_E	.04	.32	.03	−.06	1.00		
C_F	.18	−.23	.22	.16	−.06	1.00	
C_G	.48	.19	.24	.03	.02	.23	1.00

composite variable C_B, those on C_B, C_E and C_G are negligibly low. In order to get the desired score for each patient one has to reduce first of all the raw anthropometric values to standard scores. Then the average standard score for each patient on each factor should be found. It seems to save much of the unnecessary computational labor by leaving out the variables C_B, C_E and C_G and by using those in C_A, C_C, C_D and C_F only. The final result would not have been very different, even if these variables had been included. In Table 10 we present the average weighted standard scores of the anthropometric values in this study. We used the composite standard scores, without weighting, for factor F, for each psychotic group, on the ground that we had left out the composite variable for factor B, which carries a large negative weight on factor F. We nevertheless have reasons for believing that the

TABLE 9

Matrix of regression weights

	C_A	C_B	C_C	C_D	C_E	C_F	C_G
Linear factor A	.52	.07	−.16	−.25	.09	.09	.14
Magnitude factor B	.07	.55	−.07	.12	−.08	−.02	.32
Lateral factor C	−.08	.28	.56	−.10	.12	.13	−.04
Circumferential factor D	.03	−.02	.00	.61	.02	.28	−.11
Palpebral breadth E	−.35	.39	−.20	−.14	.38	.18	−.07
Size of head F	.26	−.50	.03	−.11	.07	.36	.13
G	−.25	−.39	.16	.04	.11	−.02	.48

TABLE 10

Weighted standard scores of each psychotic group

FACTOR	A		C		D		F*	
	M	$\Sigma x^{2\#}$	M	Σx^2	M	Σx^2	M	Σx^2
Non-paranoid schizophrenic	.23	8.88	−.21	2.71	−.35	7.17	−.12	15.28
Paranoid schizophrenic	.02	5.75	−.03	8.47	.37	16.71	.06	16.06
Manic-depressive	−.05	4.42	−.12	5.36	.14	10.31	−.04	11.84

* Values under this column are non-weighted composite standard scores.

\# x^2 = difference obtained by subtracting the group mean from the individual anthropometric values, squared.

standard score alone in this case was fully justified for comparing the significance of the difference in the means.

It remains to inquire as to the statistical significance of differences between the means of the various groups. We have tested this for each of the twelve pairs, using Fisher's *t* test (14). Of these differences, the only two to reach the 1 per cent significance level are those between the non-paranoid schizophrenics and each of the other two groups for the factor of trunk circumference. The score for this factor is negative for the non-paranoids, and positive for the other two groups. This confirms the conclusions earlier reached by Connolly.

One other difference, that between the non-paranoid schizophrenics and the manic-depressives in the linear factor, is significant at about the 3 per cent level.

None of the differences in the cephalic factor of any pair of groups is significant. As a matter of fact the *t* values are so low that one cannot make the inference that these groups are composed of three different populations with reference to this factor. Therefore the cephalic factor cannot differentiate one group of psychotics from another in our sample. This seems to imply that the cephalic factor is entirely different from the other physique factors; although they have some kind of antagonistic relationship, nevertheless they do not go to-

gether to form one underlying order to make differential classification of types of essential psychoses possible.

Finally, it would be worth knowing to what extent the factors we found would differentiate themselves within the same psychotic group. We left out the factor F because we had not worked out the weighted standard scores for this factor. First, we found that the *t* value for the difference between the factor A and factor C in the non-paranoid schizophrenics was 4.49. This is significant at less than the 1 per cent level. The difference between factor A and factor D is still more significant within this psychotic group. There is, however, no significant difference between the anthropometric values on factor C and factor D. This means to say that among the non-paranoid schizophrenics their linear anthropometric values exceed the mean for their group relatively more than either their lateral muscular values or the values of the thickness of the trunk. The probability of the difference between the factor A and factor D among the paranoids, however, is from 10 to 5 out of 100 that it would be not different from zero. Therefore, there is no significant superiority of thickness of the trunk to length of body in the paranoid group. However, the value of the thickness of the trunk of this group is significantly higher than that of the lateral muscular factor. This implies that factor D is probably of more importance in differentiating the paranoid type of physique than factor C. Among the manic-depressive group both the lateral muscular value and the value of thickness of the trunk are higher than the linear value, but the probability is between 30 to 20 per cent and between 5 to 2 per cent that these values would be zeros, respectively. The three significant differences so far found clearly indicate the possibility of converting these values into some sort of ratio or index which could be made use of to diagnose differentially the physique of a paranoid and a non-paranoid schizophrenic. Finally, it should be pointed out that paranoids probably form an entirely different category of mental illness and perhaps should not be confused with the other schizophrenics, as we have seen the paranoids' physique is different from that of the remaining types of schizophrenic patients.

SUMMARY

A factorial analysis of Thurstone's techniques was applied to the anthropometric data obtained by Connolly. Seven factors were isolated, five of which yielded to meaningful interpretation. The factors identi-

fied are: (1) a linear factor, (2) a magnitude factor, (3) a lateral-musclar factor of extremities, (4) a thickness of trunk factor, and (5) a cephalic factor. These factors are slightly correlated. The cephalic factor correlates negatively with factors (1), (3), and (4). Regression weights were obtained for the four most significant factors and the raw anthropometric values of 96 male psychotic patients were converted into standard scores multiplied by the respective regression weights. Fisher's *t* values were then found to test the significance of the differences in these weighted scores on the factors. The thickness of trunk factor was found to differentiate significantly the non-paranoid schizophrenics from the paranoid schizophrenics and the non-paranoid schizophrenics from the manic-depressives. The linear values in non-paranoid schizophrenics is significantly higher than that of the lateral muscular value and the trunk-thickness value. In the paranoid group the trunk-thickness value is significantly higher than that of the lateral muscular value. The analytical result implies that paranoids and non-paranoid schizophrenics really belong to different categories.

The factors found have no doubt a true physiological meaning. There are a number of factors that have to do with the growth of the organism. To obtain further light on the nature of these factors and perhaps their endocrinological significance, we should use anthropological measurements of a large random sample of a group of individuals drawn from a homogeneous group belonging to a specific race. Were this done light of considerable value might be thrown on the biological factors of human growth.

Our present study did not envisage this important problem, but was concerned with an attempt to determine whether or not some statistically significant differences of physique characterize patients suffering from different psychotic conditions. Our study leads to the conclusion that such statistically significant differences are found and suggests a technique for the investigation of the physiology of growth.

BIBLIOGRAPHY

1. ANASTASI, A. Further studies on the memory factor. *Arch. of Psychol., N. Y.*, 1932, *142*, p. 40.
2. BURT, C. The analysis of temperament. *Brit. J. Medical Psychol.*, 1938, *17*, pp. 158-188.
3. ————. The Factors of the Mind. Macmillan, *N. Y.*, 1941, p. 249.
4. ————. Factor analysis of physical growth. *Nature*, 1943, *152*, p. 75.

5. CARPENTER, A. An analysis of the relationships of the factors of velocity, strength, and dead weight to athletic performance. *Research Quarterly*, 1941, *12*, 1, pp. 34-39.
6. ————. An anthropometric study of masculinity and femininity of body build. *Research Quarterly*, 1941, *12*, 4, p. 714.
7. COHEN, J. Determinants of physique. *J. of Mental Science*, 1938, *84*, pp. 495-512.
8. COHEN, J. I. Physique size and proportions. *Brit. J. Medical Psychol.*, 1939-41, *18*, pp. 323-337.
9. ————. Are differences in size between parts of the body due to general or specific factors? *Proc. Nat. Acad. Science, Washington*, 1940, *26*, pp. 524-526.
10. COLEMAN, J. W. Pure speed as a positive factor in some track and field events. *Research Quarterly*, 1940, *11*, 2, pp. 47-53.
11. CONNOLLY, C. J. Physique in relation to psychosis. *Studies in Psychol. and Psychiatry*, 1939, IV, 5, Catholic Univ. of Amer. Press.
12. COOMBS, C. H. A criterion for significant common factor variance. *Psychometrika*, 1941, *6*, 4, pp. 267-272.
13. DEARBORN, W. F. and ROTHNEY, J. W. M. Predicting the Child's Development. Sci.-Art Publishers, *Cambridge, Mass.*, 1941, p. 270.
14. FISHER, R. A. Statistical Methods for Research Workers. Oliver & Boyd, *London*, 1938.
15. HAFSTÄTTER, P. R. Ueber Typenanalyse. *Arch. ges. Psychol.*, 1940, *105*, pp. 305-403. *Psychol. Abst.*, 1941, No. 3891.
16. HAMMOND, W. H. An application of Burt's multiple general factor analysis of the delineation of physical types. *Man*, Nos. 1-18, 1942, Jan.-Feb.
17. HOLZINGER, K. J. and HARMAN, H. H. Factor Analysis. Univ. of Chicago Press, 1941.
18. HUXLEY, J. S. Problems of Relative Growth. Dial Press, *N. Y.*, 1932, p. 154, p. 165.
19. MCCLOY, C. H. An analysis for multiple factors of physical growth of different age levels. *Child Develop.*, 1940, *11*, pp. 249-277.
20. MCDONOUGH, SR. M. R. Character and body build in children. *In:* Aspects of the New Schol. Philosophy, 1932, Benziger Bro., *N. Y.*, pp. 236-242.
21. MONAGHAN, E. A. Major factors in cognition. *Stud. in Psychol. and Psychiatry*, Vol. III, No. 5, 1935.
22. MULLEN, F. Factors in the growth of girls seven to seventeen years of age. Univ. of Chicago, Ph.D. dissertation, 1939.
23. REES, W. L. Physical and psychological aspects of constitution. *Eugenics Rev.*, 1945, *37*, pp. 23-27.
24. ————, and EYSENCK, H. J. A factorial study of some morphological and psychological aspects of human constitution. *J. of Mental Science*, 1945, *91*, pp. 8-21.
25. SANFORD, R. N., *et al.* Physique, personality and scholarship. *Monog. Soc. for Res. in Child Develop.*, 1943, VIII, *34*, 1, p. 57.
26. STERRETT, J. E. The prediction of heart size by anthropometric measurements in adult males. *Research Quarterly*, 1935, *7*, 3, p. 112.

27. SPEARMAN, C. The Abilities of Man. Macmillan, *N. Y.*, 1927, pp. 140-145.
28. THURSTONE, L. L. The Vectors of Mind. Univ. of Chicago Press, 1935.
29. ————. A new rotational method in factor analysis. *Psychometrika*, 1938, *3*, 4, pp. 199-218.
30. ————. Factor analysis and body types. *The Psychometric Lab.*, No. 24, September, Univ. of Chicago, 1945, pp. 1-7.
31. TRYON, ROBERT C. Cluster Analysis. Edwards Bro., *Ann Arbor, Mich.*, 1939.
32. TUCKER, L. R. The role of correlated factors in factor analysis. *Psychometrika*, 1940, Vol. 2, pp. 121-136.
33. WALDROP, R. S. A factorial study of the components of body-build. *Psychol. Bulletin*, 1940, *37*, p. 578.
34. WISSLER, CLARK. The correlation of mental and physical tests. *Psychol. Rev., Monthly Supplement*, 1901, *3*, 6 (16), pp. 33, 61.
35. WOLFLE, D. Factor Analysis to 1940. Univ. of Chicago Press, 1940.

PROBABILITIES OF DEATH IN CLOSED POPULATION GROUPS
ILLUSTRATED BY PROBABILITIES OF DEATH OF WHITE FATHERS AFTER BIRTH OF CHILDREN

BY T. J. WOOFTER
Director of Research, Federal Security Agency

MANY demographic problems require the estimate of survival rates or their complement, probabilities of death, for a number of years after a date when a particular status is acquired. Among such groups are those composed of fathers, mothers, married persons, new entrants to the labor market, immigrants, and veterans.[1] A particular age group in an enumerated population also acquires such status at the time of enumeration. The following discussion evaluates four methods of calculating probabilities of death in closed populations.

The characteristics of such closed groups which make them subject to estimate of average probabilities of death are:

1. The individuals have acquired a status at a particular time and will remain in that status until death.

2. The initial average age of individuals will be above 15.

3. The group is closed in the sense that no new entrants are received. If the ages of annual increments to the group are known, each annual "crop" may be treated as a separate closed group.

The data used in this article to illustrate the problem refer to white fathers in the years 1927 to 1944 (i.e., fathers of children under 18 years of age in 1945). Fathers in each calendar year were considered as a separate closed group. Their ages were estimated by averaging the ages of fathers for three years during the period, and their mortality rates were assumed to be those of the U. S. Life Table, White Males, 1939-41.

[1] Pioneer work was done on this problem in calculating probability of orphanhood by A. J. Lotka and discussed in "Orphanhood in Relation to Demographic Factors," *Metron*, Vol. 9, No. 3.

The age distribution of fathers who survive up to successive anniversaries of the birth of a child form a pattern of related annual frequency distributions, each one differing from that of the previous year. This difference arises from the fact that after the early years of life, mortality rates increase steadily. The result is that the number of deaths above the average age is constantly larger than that below the average. Furthermore, as every surviving individual in the group ages one year, the average age does not advance a full year because a smaller proportion of the older members survive. (See Table A, Fig. A.)

It should also be noted that the age at which the mean mortality rate occurs differs from the mean age. In most groupings over a

TABLE A

Age of fathers at birth of children (1927-1944) and of those surviving 40 years and 60 years; U. S. Life Table, 1939-41, White Males

POPULATION (P_x) AT BIRTH OF CHILD		DEATH RATE	P_{x+40} SURVIVING 40 YEARS		P_{x+60} SURVIVING 60 YEARS	
Age	Number living	q_x	Age	Number living[1]	Age	Number living[2]
17½	1 520	.00179	57½	1 175	77½	417
22½	19 310	.00235	62½	13 308	82½	2 935
27½	28 300	.00255	67½	16 442	87½	1 664
32½	22 120	.00314	72½	9 949	92½	331
37½	14 900	.00429	77½	4 500	102½	30
42½	8 070	.00625	82½	1 311		0
47½	3 660	.00942	87½	235		0
52½	1 350	.01418	92½	23		0
60	770	.02548	100	1		0
Total	100 000			46 944		5 377

[1] Computed by formula P_{x+40}/P_x (modified long method)

[2] Computed by formula P_{x+60}/P_x (modified long method)

broader span than five years, the age of mean mortality is older than the mean age. In narrow groupings the principal exceptions to this relationship will be found in the youngest sections in the life table or in groups heavily skewed toward the lower ages. In these circumstances

the initial age of mean mortality may be slightly lower than the mean age and may actually increase more than one year between two anniversaries, although the mean age increases less than one year.

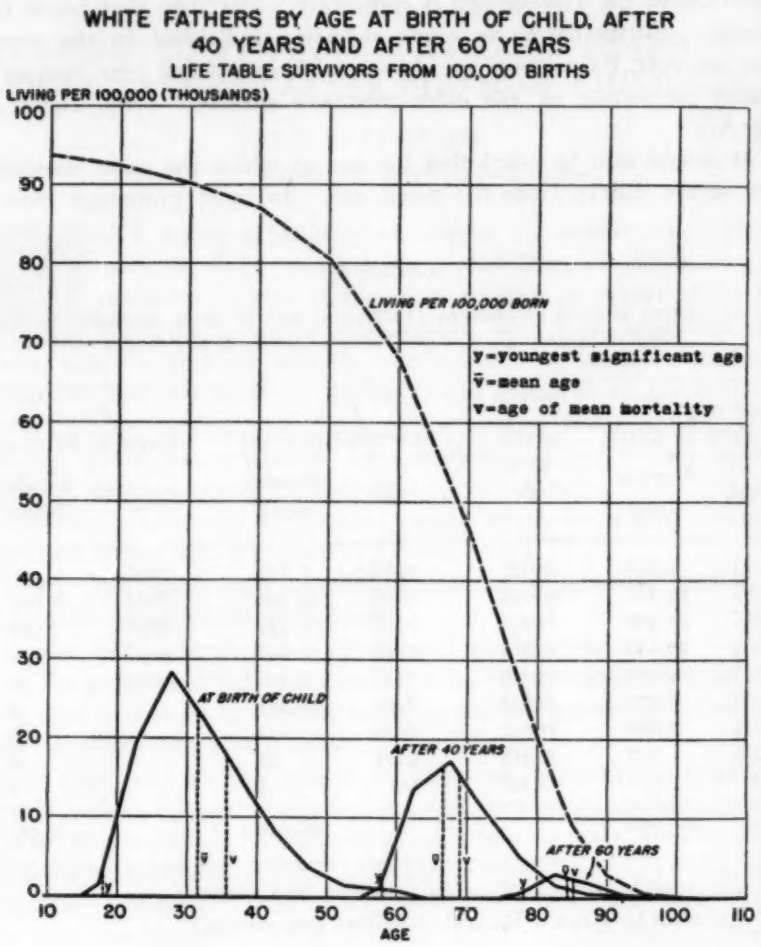

FIG. A. AGES OF SURVIVING FATHERS FROM TABLE A. SURVIVORS FROM BIRTH, U. S. L. T. 1939-41

The following notation is used in the computations:
Referring to the closed frequency distribution—

x = midyear of an age step when status is acquired
P_x = population acquiring status in age step with midpoint x
ΣP_x = total population in all ages acquiring status
a = anniversary of acquisition of status
q_x = mortality rate in year beginning with age x
$\bar{q}^a$ = mean mortality rate for year beginning with anniversary a
y = approximate age in years of youngest individual
v = age of mean mortality at the beginning of year when status is acquired (life table age corresponding to $\bar{q}^0$)
Q^a = probability of an original member of group dying during the year beginning with anniversary (a) of acquisition of status (mean probability of death)

$\sum_0^{a-1} Q^a$ = probability of an original member of group dying between anniversary 0 and anniversary a.

In the life table—

$q_v = Q^0 = q^0$
ω = last age in the life table which an individual will probably attain, but not survive
k_a = the increment in the age of mean mortality added to v for an annual increment of a

$\sum_0^{a-1} k_a$ = years added to age v to obtain increase in the mean mortality age from anniversary 0 to anniversary a

$\bar{k}$ = mean increment in age of mean mortality added to v for values of a from 0 to $\omega - y$

k'_a, $\Sigma k'_a$ and $\bar{k}'$ = increments in age of mean survival.

I. EXACT LONG METHOD

1. The exact, but sometimes laborious, method of calculating probabilities of death in closed groups for each year after the acquisition of status is to compute the probability of death at each single age and add the number dying at each age in order to arrive at total number dying in the whole distribution.

2. If ages are grouped in the frequency distribution which approximates normal, it is permissible to assume an equal number of individuals in each of the initial ages of a frequency step and calculate the probability of death up to the first anniversary within each step by the formula:[2]

$$Q^0 = 1 - \frac{\frac{L_1}{L_0} + \frac{L_2}{L_1} + \frac{L_3}{L_2} + \frac{L_4}{L_3} + \frac{L_5}{L_4}}{5}$$

and to repeat this calculation up to each anniversary in the period of prediction.

Again, the probability in all steps is added to arrive at the total probability of death in the whole distribution.[3]

II. MODIFIED LONG METHOD

A somewhat shorter, but less exact,[4] method is sometimes used.

---

[2] L_x being the life table value approximating the number living in the middle of the year (stationary population).

[3] A special application of the method is adapted to the estimate of survival of closed groups in the population above a specified age (65 years of age and over, 75 and over, etc.). From the long method, the mean mortality rate

$$\bar{q}^x = \frac{\Sigma l_x \, q_x}{\Sigma l_x}, \text{ but}$$

$$\Sigma l_x q_x = \Sigma d_x, \text{ and}$$

$$\Sigma d_{w+1} = l_w$$

$$q^x = \frac{l_w}{\Sigma l_{w+1}} \left(\text{ or } \frac{L_w}{T_w} \right).$$

For the second year

$$q^{x+1} = \frac{l_w}{\Sigma l_{w+1}} \left(\text{ or } \frac{L_w}{T_w} \right).$$

This method, of course, assumes that the population group projected has an age distribution approximating that of a stationary population which is usually fairly accurate at the older ages unless they have been severely skewed by past migration or sharp fluctuation in annual number of births.

[4] When groupings as narrow as five ages are used, the error is negligible in short range projections for age distributions in which v is near the middle of the life span. The error increases in proportion to the width of the age group and the extent to which the older ages predominate in the distribution. (*Cf.* previous discussion of the relationship between the mean age and the age of mean mortality. The close correspondence of results obtained from (Table E) cal-

It assumes that the midpoint of each step may be used as a location of the average death rate and computes the probability of death within each step by multiplying the death rate at the midpoint age (q_x) by the number living (P_x) in that age step to obtain the number of deaths in that age step. This process must be continued for each year, subtracting the number dying during the year beginning at age x to obtain the number surviving at the beginning of the age $x+1$ and repeating the multiplication as shown in Table B. When each frequency step

TABLE B

Computation of modified long method illustrated by mortality of fathers age 30-34 at birth of child

AGE	DEATH RATE	NUMBER LIVING (P_x)	DEATHS
32½	.00314	22 120	69.5
33½	333	22 051	73.2
34½	352	21 978	77.4
35½	375	21 901	82.1

has been treated in this manner, the total number of deaths in the whole group in a year is obtained by summing deaths for that year in all frequency steps. Then the probability of death in that year for the original group is the summed deaths divided by the original number in the group. Calculation by either the exact long method or the modified long method can be protracted if the probabilities are to be estimated for a long series of years or if it is desired to subdivide the result by repeating the calculation for such classes as sex, race, or economic status.

III. SHORT METHODS

It would, therefore, be useful to have a short method which would reduce the length of the computation and at the same time produce reliable results.

culating probabilities of death of white fathers by the exact and modified methods is due to the compensating effects of the inclusion of some young groups whose initial ages of mean mortality are below the mean ages and whose mean mortality ages increase more than one year per anniversary, together with older groups whose trend in mean mortality in relation to mean age is opposite.)

Method A

Such a method is one by which it is assumed that all the indviiduals in a group are initially concentrated at the age of average mortality (v)—i.e., where q_v represents the probability of death in the first year of exposure $Q^0 = q_v$. If, thereafter, subsequent increases in the age of average mortality can be computed, average mortality rates at these ages can be read directly from the appropriate life table and applied to a cohort beginning at age v.

1. The first step is to locate the age v. This is done by actually calculating the mean mortality rate for the first year of exposure by the long method and locating corresponding age in the life table by inspection and interpolation. From the age distribution in column 2 of Table A and the mortality rates of U. S. white males, 1939-41 (see column 3, Table A), the mean mortality is:

$$\bar{q}^0 = \frac{\Sigma P_x q_x}{\Sigma P_x} = .003775 = q_v,$$

or v is located at age 35.60 (by inspection).

2. If the annual rates of increase of the ages of mean mortality are k_a, the subsequent ages of mean mortality are given by $v + \Sigma k_a$.

The mean value of k_a ($\bar{k}$) over the entire life span may be derived as follows:

y equals youngest age in the distribution, and ω equals the oldest age which any member of the group may attain but probably not survive.

The members of the group who are youngest when the status is acquired have the maximum expectation of life; hence the last surviving member will probably be one who acquired the status at approximately age y. His maximum life span will be $\omega - y$ years, after which time he will probably be the last survivor of the group and his probability of death will become the average probability for the surviving members—i.e., after the lapse of $\omega - y$ years, the age of mean mortality increases by $\omega - v$ years.

Therefore, when:

$$a = \omega - y$$

$$\Sigma k_a = a\bar{k} = \omega - v, \text{ and then}$$

$$\bar{k} = \frac{\omega - v}{\omega - y}$$

Illustration: In life table used—

$$\omega = 108$$

From frequency distribution, Table A—

$$y = 18$$

$$v = 35.6$$

$$\bar{k} = \frac{108 - 35.6}{108 - 18} = .804$$

But $\bar{k}$ cannot be relied on as a constant prediction of k_a. The most convenient method of approximating k_a for various values of a is as follows:

3. k_0 (advance in age of mean mortality from the first to the second year of exposure — anniversary 0 to anniversary 1) may be determined by calculating by the long method the mean rate of mortality for the second year as well as for the first year and comparing it with the first and determining by inspection the fraction of a life table year represented by the difference.

Illustration: $k_0 = \dfrac{\bar{q}^1 - \bar{q}^0}{q_{v+1} - q_v}$

$$k_0 = \frac{3992 - 3775}{4032 - 3775} = .844$$

4. It may be determined empirically that approximation to the ages of mean mortality may be obtained by a parabola[5] plotted from the mid-ordinate, y_0, $x = a - \dfrac{\omega - y}{2}$,

$$\Sigma k_a = y_0 + bx - cx^2$$

The necessary constants are:

$$y_0 = \frac{\omega - y}{2}\left(\frac{k_0 + \bar{k}}{2}\right) \pm \bar{k}$$

$$b = \bar{k}$$

$$c = \frac{y_0 - \left(\dfrac{\omega - v}{2}\right)}{\left(\dfrac{\omega - y}{2}\right)^2}$$

[5] A somewhat simpler approximation of k_a yielding results within 2 per cent of accuracy in most distributions is obtained from linear interpolation between k_a and $\bar{k}$ (.844 and .804), yielding an annual increment in k_a of $(.804 - .844)/45 = -.0009$.

The sign of $\pm\bar{k}$ in the expression for y_0 agrees with the sign of $k_0 - \bar{k}$.

Substitution from illustrations of Steps 2 and 3 to obtain these constants:

$$y_0 = 45(.804 + \frac{.844 - .804}{2}) \pm .804 = 45(.824) + .804 = 37.88$$

$$b = \bar{k} = .804$$

$$c = \frac{37.88 - 36.20}{(45)^2} = \frac{1.68}{2025} = .0008$$

Substituting to obtain the age of mean mortality at the 17th anniversary:

$$x = 17 - 45 = -28$$

$$\Sigma k_a = 37.88 - 28(.804) - (28)^2\ (.0008)$$

$$= 37.88 - 22.51 - .63 = 14.74$$

$$v + \Sigma k_a = 35.6 + 14.74 = 50.3$$

5. Finally, values of $q_{v+\Sigma k_a}$ when so determined are used to determine the probabilities of death Q^a by starting with $P_v = 1$ and multiplying by q_v to obtain Q^0, reducing the survival ratio each year by subtracting Q^a from $P_{v+\Sigma k_a}$ to obtain $P_{v+\Sigma k_{a+1}}$ (*Cf.* Table C).

Method B

When only mortality rates are available, the computation is limited to Method A. But if tables with precalculated l_x values are available, an alternate method of computation of Q^a is more convenient, especially for computing cumulated mortality up to anniversary a, $\sum_{0}^{a-1} Q^a$, without reference to intervening single year mortality rates. This method is based on the relationship:

TABLE C

Probability of death of white fathers from anniversary 0 through year beginning with anniversary 17. Short method A. Ages of fathers 1927-1944; mortality rates white males 1939-41

ANNIVERSARY a	MEAN MORTALITY AGE $v + \Sigma k_a$	MEAN MORTALITY RATE $q_{v+\Sigma k_a}$	PROBABILITY OF DEATH WITHIN THE YEAR Q^a	PROBABILITY OF DEATH CUMULATED[1]	MODIFIED LONG METHOD CUMULATED
0	35.6	.00378	.00378	.00378	.00378
1	36.4	398	396	774	775
2	37.3	423	420	1194	1195
3	38.2	450	445	1639	1638
4	39.0	476	468	2107	2108
5	39.9	509	497	2604	2606
6	40.8	546	531	3135	3135
7	41.6	582	564	3698	3697
8	42.5	625	603	4302	4295
9	43.4	673	644	4948	4931
10	44.2	718	683	5631	5611
11	45.1	773	730	6361	6336
12	45.9	826	773	7134	7108
13	46.8	889	826	7960	7925
14	47.6	950	875	8835	8807
15	48.5	1022	932	9767	9740
16	49.3	1095	986	10753	10731
17	50.2	1175	1048	11801	11787

[1] Probability of dying between anniversary 0 and anniversary $a+1$.

$$\frac{1}{l_x} \cdot \sum_{0}^{a-1} d_x = 1 - \frac{l_{x+a}}{l_x}$$

or in the short method:

$$\sum_{0}^{a-1} Q^a = 1 - \frac{l_{v+\Sigma k_a}}{l_v}$$

Whereas Method A is computed from graduated rates of mean mortality, Method B is computed from graduated rates of mean survival. $\bar{k}$ in Method B is the same value as $\bar{k}$ in Method A.

$k'_0 = 1$, and $\Sigma k'_a$ may be determined by linear interpolation [6] or from a parabola with the constants—

$$y_0 = \frac{\omega - y}{2}\left(\frac{1 + \bar{k}}{2}\right) \pm \bar{k}$$

$$b = \bar{k}$$

$$c = \frac{y_0 - \left(\frac{\omega - v}{2} \pm 1\right)}{\left(\frac{\omega - y}{2}\right)^2}$$

The signs of $\pm \bar{k}$ and ± 1 agree with the sign of $k_0 - \bar{k}$ as determined in Method A.

(*Cf.* Table D for computation by this method.)

TABLE D

Method B — Results for selected ages of cumulated probability of death up to anniversary a (data from white fathers and Life Table, White Males 1939-41)

ANNIVERSARY	$v + \Sigma k_a$	$l_{v+\Sigma k'_a}$	CUMULATED PROBABILITY OF DEATH	
			$1 - \frac{l_{v+\Sigma k_a}}{l_v}$	Exact method
0	35.6	88 519	0	0
5	40.5	86 657	.02108	.02108
10	45.3	84 091	.04907	.04931
18	52.7	77 845	.12059	.11919
40	72.1 [1]	41 204 [2]	.53464 [2]	.53094
60	87.8	4 943	.94410	.94351

[1] $\Sigma k'_a = (41.39 + .804)\ (-.00207x)$

$$\sum_0^{\omega} k'_a = (41.39 - 6)\ (.804 - 36)\ .00207 = 41.39 - 4.89 = 36.50$$

$$v + \sum_0^{\omega} k'_a = 35.6 + 36.5 = 72.1$$

[2] $1 - (41204/88519) = 1 - .46536 = .53464$

[6] As in Footnote 5.

It will be noted that for a single projection of cumulated probabilities, Method B is not much shorter than the alternate long method suggested in 1 (2). The saving is in proportion to the number of steps in the frequency distribution and the number of separate estimates of cumulated probability which are required by the problem in hand.

Obviously, the cumulated probabilities for the 18 separate years shown in column 4, Table C, can be as conveniently calculated by Method B as by Method A.

The relationship between Short Methods A and B will be more apparent if we consider the relationship:

$$Q_a = q_v + \Sigma k_a \left(\frac{l_{v+\Sigma k_a - 1}}{l_v} \right)$$

i.e., $P_{v+\Sigma k_a}$ as computed by Step 6, Method A, approximates $\frac{l_{v+\Sigma k_a - 1}}{l_v}$ as read from the life table ages determined by Method B.

Table E compares the results of the four methods discussed.

TABLE E

Comparison of cumulated probabilities of death up to years beginning with specified anniversaries, computed by four methods

ANNIVERSARY	METHOD			
	Exact long	Modified long	Short A	Short B
18	.11919	.11787	.11801	.12059
40	.53094	.53056	.52343	.53464
60	.94351	.94623	.93976	.94414

SUMMARY

Mathematical principles

The short methods proposed in this paper are based on the following principles: That a closed group survives until the youngest member (initial age y) reaches the limiting age in the life table (age ω); that in these $\omega - y$ years the age of mean mortality advances from the age of mean mortality when status is acquired (v) to ω, i.e., when $\omega - y$

anniversaries have elapsed after the acquisition of status, the age of mean mortality has advanced $\omega - v$ years. Thus, the age of mean mortality of the group approaches the age of the youngest member as the age of the youngest member approaches ω. Therefore, if the approach of the ages $v + \Sigma k_a$ to ages $y + a$ is estimated as in the short methods presented herewith, the average mortality experience of the group may be read from the functions corresponding to these ages in the appropriate life table.

Practical applications

In computing the survival (or probabilities of death) in a closed status group, such as fathers, mothers, immigrants, veterans, etc., which has a known age distribution and known age specific mortality rates, the mean experience of the group may be calculated more conveniently by the short methods proposed in this paper than by any other accurate method. The accuracy of such computations by either short method A or B is greater for long-time projections than computations made by any other method, except the exact long calculation of the probabilities of death in each single year of age. In groups with the initial age of mean mortality near the middle of the life span, computations by the modified long method, using frequency steps as narrow as five ages, are for all practical purposes as accurate as the exact long method for short-run predictions. When, however, older ages predominate in the distribution, or age groups are broader than five years, or the period of prediction is lengthened, the short methods tend to be more accurate than the modified long method. The short methods are also adapted to the calculation of survivors of broad and irregularly distributed age groups in an enumerated population, such as the total male working population of a state which has been disturbed by heavy migration at various periods. If, however, it may be assumed that the age distribution of an enumerated population conforms reasonably closely to that of a stationary population, then special methods beyond the scope of this paper are applicable.